

Designing Transparent Teacher Evaluation: The Role of Oversight Panels for Professional Accountability

JENNIFER GOLDSTEIN

City University of New York, Baruch College

Background/Context: *Peer Assistance and Review (PAR), or peer review as it has historically been called, has existed in a handful of school districts since the early 1980s. In 1999, California became the first state to pass PAR legislation; at that time, a major district had not implemented the policy in over a decade.*

Setting: *This is an in-depth study of one urban district in California, given the pseudonym Rosemont, as it implemented PAR following California's legislation.*

Program: *With PAR, designated "PAR coaches"—teachers identified for excellence and released from teaching duties full-time for 2–3 years—provide mentoring to teachers new to the district or the profession, intervention for identified veteran teachers experiencing difficulty, and the formal personnel evaluations of both groups. These PAR coaches are not school based but rather report to an oversight panel composed of teachers and administrators from across the district.*

Research Questions: *A companion study previously found that the rate of dismissals increased dramatically after the implementation of PAR in Rosemont. This study examines one aspect of Rosemont's PAR program, its oversight panel. The study examines three questions that in turn address the design, process, and outcomes of PAR and the PAR panel in Rosemont: (1) How did the PAR panel work? (2) How, if at all, did the presence of an oversight panel affect the teacher evaluation process? (3) How, if at all, did the presence of an oversight panel affect personnel outcomes?*

Participants: *The study involved all PAR coaches (10) and all members of the PAR panel (9) for the first year of program implementation.*

Research Design: *The study employed an embedded single-case design in Rosemont over 1 1/2 years.*

Data Collection and Analysis: *The study relied on observations (311 hours of meetings)*

and semistructured interviews (39). I used QSR NVivo data management. I coded observation scripts and interview transcripts were coded to a schema developed from the progressive coding patterns that I observed, and analyzed them in response to the research questions.

Findings: The Rosemont data demonstrate that the PAR panel both supported the coaches to do their jobs and held them accountable. The community of educators created by PAR and the PAR panel appears to have proved a more rigorous, evidence-based check on classroom teaching performance. I also present data demonstrating the shortcomings of the PAR panel structure in the face of enduring norms against accountability in education.

Conclusions: Significant shifts in organizational structure occurred to support the PAR coaches in their role as evaluators. Making the typically solitary practice of teacher evaluation transparent to colleagues fundamentally altered the nature of educational accountability.

trans • pār • ent, *a.* [Fr., from L. *trans*, through, and *parens* (-*entis*), ppr. of *parere*, to appear.]

1. transmitting light rays so that objects on the other side may be distinctly seen; capable of being seen through; as, window glass is *transparent*: opposed to *opaque*. . .

— *Webster's Unabridged Dictionary*

Instructional leadership is often defined as creating and maintaining the conditions for high-quality teaching (Murphy & Seashore Louis, 1999; Spillane, Halverson, & Diamond, 2001). A primary goal of instructional leaders, therefore, is to design organizational structures that promote high-quality teaching (Halverson, 2003). This article explores one such structure designed to improve the quality of teaching by improving the quality of teacher evaluation. It examines a Peer Assistance and Review (PAR) program, specifically one aspect of the program—its oversight panel—and asks how an oversight panel alters the practice of teacher evaluation. The core argument of the article is that oversight panels have the potential to alter the transparency of the teacher evaluation process fundamentally, and in turn the nature of accountability.

Practitioners and researchers alike bemoan the ineffectiveness of teacher evaluation. Various factors are blamed, with principals' lack of time and teacher unions' defense of ineffective teachers usually topping the list (Painter, 2000). Other problems, however, include evaluators' lack of training in conducting evaluations (Loup, Garland, Ellett, & Rugutt, 1996; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984), a lack of grade-level or subject match between the evaluator and the teacher being evaluated (Darling-Hammond, 1984), principals' tendency to avoid conflict (Bridges, 1986), evaluation instruments that typically

apply the same standards and criteria to all teachers regardless of years of experience (Koppich, 1998), and evaluation instruments that produce very little actual data (Hunter, 1988; Loup et al.). Notably, teachers rarely receive negative evaluations despite widespread discontent with the quality of teaching practice (Tucker, 1997). Instead, the notorious “dance of the lemons” (Bridges, 1992) is pervasive, whereby low-performing teachers are passed from school to school repeatedly.

What is remarkable about the teacher evaluation program presented in this article, therefore, is that district leaders—both administrators and teacher union leaders—designed a new structure to alter the shape of teacher evaluation radically in the district.¹ Rather than leave principals to conduct teacher evaluation alone in isolation, and in opposition to the teacher union, some school districts have taken a more coordinated and comprehensive approach to the evaluation of teachers. With PAR, designated “PAR coaches”—teachers identified for excellence and released from teaching duties full-time for 2–3 years—provide mentoring to teachers new to the district or the profession, intervention for identified veteran teachers experiencing difficulty, and formal personnel evaluations of both groups. These PAR coaches are not school based but rather report to an oversight panel composed of teachers and administrators from across the district. PAR shows promise as a teacher evaluation system (Darling-Hammond, 1984; Gallagher, Lanier, & Kerchner, 1993; Goldstein, 2003; Hewitt, 2000), and prior research suggests that the oversight panel is at the center of PAR’s effectiveness (Goldstein, 2007b). Specifically, the oversight panel generates a transparent teacher evaluation process that allows the coaches to be supported and held accountable for their evaluative decisions.

This article is drawn from a multiyear, in-depth study of one urban school district in California as it implemented PAR, given the pseudonym Rosemont Unified School District. Because PAR has potential as an improved mechanism for teacher evaluation, it is important to understand the role that the panel played in the policy’s outcomes. This study therefore examines three questions that in turn address the design, process, and outcomes of PAR and the PAR panel in Rosemont: (1) How did the PAR panel work? (2) How, if at all, did the presence of an oversight panel affect the teacher evaluation process? (3) How, if at all, did the presence of an oversight panel affect personnel outcomes?

The act of needing to defend one’s decisions with evidence naturally alters and guides the decision-making process. One key finding of prior research (Goldstein, 2007b) is that 10%–12.5% of beginning teachers and almost all veterans across the first 4 years of the PAR program in Rosemont were removed from classroom teaching, a stark figure

compared with education's norm, which one study placed at 0.1% (Tucker, 1997). Regardless of one's reaction to fairly unprecedented rates of teacher firings, there was clearly something unusual at play. What accounts for the phenomenon? It is entirely possible that without PAR panel oversight, the coaches might not have dismissed teachers at such an unusually high rate; similarly, if principals had to defend their evaluative decisions to such a body, higher rates of dismissals might occur with traditional teacher evaluation.

In other words, it certainly cannot be argued that replacing evaluation by a principal with evaluation by a PAR coach led to increased dismissals in Rosemont, but rather that significant shifts in organizational structure occurred to support the PAR coaches in their role as evaluators. Some of these structural shifts are discussed in companion papers. Time plays an important role (Goldstein, 2007b) because the coaches were engaged in teacher evaluation full time, whereas principals are stretched thin and have limited time to bring to the task. Having the same person conduct both formative and summative assessment with PAR, thereby formally linking professional development and evaluation, is another significant structural shift (Goldstein, 2007a). Putting district administrators and the teacher union on the same side of the table in the evaluation process radically changes the nature of the conversation (Goldstein, *in press*). The argument presented in this article is that making the typically solitary practice of teacher evaluation transparent to colleagues fundamentally alters the nature of educational accountability.

In the following section, I outline the study's design and methods. I then present Rosemont case data to describe the workings of the PAR oversight panel and to elucidate its effects on teacher evaluation with PAR. The Discussion section, which follows the findings, examines how the Rosemont PAR panel created a teacher evaluation process that was more transparent than the norm, and how that transparency in turn affected accountability for the quality of teaching.

STUDY DESIGN AND METHODS

The study employed an embedded single-case design (Yin, 2003) of Rosemont USD. In 1999, California became the first state to pass legislation implementing PAR. Whereas many California districts and/or teacher unions shied away from the idea of PAR and focused merely on compliance with the new legislation, key leaders in Rosemont intended to implement the new policy as fully as possible, making the district a rich site in which to study PAR in California at the time.² The study examined PAR in Rosemont in-depth for a year and a half, with some follow-up data

collection occurring in the second and fourth years of the program.

Almost one year prior to program implementation, district and teacher union leaders in Rosemont approached me to “document” the PAR implementation efforts. I served in this capacity for a year and a half, in exchange for using the site for a dissertation study. I was given wide access to program participants and events but in no way acted as a participant in the process during this time. At the end of this period, in June following the first year of program implementation, I provided a written report and two presentations—one to coaches and one to panel members—highlighting notable successes and challenges from the year (many of which were wholly unrelated to the research questions addressed here or in prior companion studies). Relationships established during this time allowed me to also conduct some follow-up interviews and focus groups at the end of the second and fourth years of implementation.

Rosemont has approximately 100 schools and 3,000 teachers. The members of the PAR panel selected 10 PAR coaches for the first year of the program, who supported 91 classroom teachers (called “participating teachers” or “PTs”). The 91 PTs included 88 beginning teachers and three veteran teachers across 28 schools. Coaches had initial caseloads of nine or 10 PTs.³ The sample included the district’s nine members of the PAR oversight panel (teachers and administrators) and 10 PAR coaches. In addition, three of the 10 coaches were chosen for more in-depth data collection. This choice was influenced by their demography (years of experience, gender, and ethnicity) and degree of “sensemaking” (Spillane, Reisen, & Reimer, 2002; Weick, 1995) about the reform as observed in ongoing meetings. The sample of 19 teachers and administrators included three African American women, three Latina women, one Chinese American woman, six White women, and six White men.

The study relied primarily on observations and interviews. Data collection began with the inception of the panel in the spring prior to the first year of policy implementation. It included the selection of the coaches by the panel and the professional development activities that occurred over the summer. I then attended all panel meetings and hearings (approximately monthly) and almost every coach meeting (weekly) for one year. I scripted these meetings, a process that involves creating an approximately verbatim account of interactions that looks like the script of a play. Meetings were also tape recorded, allowing me to fill in gaps or make corrections to the script after the conclusion of meetings. In total, I observed and scripted approximately 311 hours of meetings. Please see the Appendix Table A1 for a breakdown of observation hours.

In addition to observing meetings, I conducted 39 semistructured interviews with PAR coaches and panel members during the first year of

the program. I interviewed panel members and coaches in the fall and spring, and the three case study coaches in the winter as well.⁴ Interviews lasted between one and three hours; all but three were tape recorded and transcribed. Please see the Appendix Table A2 for a breakdown of interviews conducted.

Data collection and analysis were interwoven from the outset of the study. Ongoing analysis of the data relied on summaries of field notes, analytic memoing, and coding (Miles & Huberman, 1994). I originally used the qualitative software QSR NUD*IST 4 for data management⁵ and created coding schema from the progressive coding patterns that I observed.

For a second wave of analysis, which formed the basis of this article, I converted the database to QSR NVivo and expanded the original coding schema. In particular, codes for various “distributed accountability” (Goldstein, 2007c) themes were added to the original coding tree after the panel emerged as a key factor in the findings from an earlier study (Goldstein, 2003). Transcripts of interviews and panel hearings were then recoded for the new themes. Specifically, the original coding node “distribution/sharing of role tasks and functions” was expanded to include the subset “distributed accountability.” The coding rule for distributed accountability read, “pertaining to a distribution of leadership involving holding one another accountable for practice.” The distributed accountability node was in turn broken into five subnodes addressing accountability between specific PAR roles: (1) teacher union and school district, (2) panel and coaches, (3) panel and principals, (4) coaches and principals, and (5) principals and PTs. Accountability between coaches and PTs was so central to the original study that it existed as a separate node unto itself, prior to the expansion of the coding schema. Examination of the data for PAR’s four *formal* lines of accountability—numbers 1 to 3 above, plus accountability between coaches and PTs—forms the basis of a companion paper (Goldstein, 2007c). However, given prior findings suggesting the central importance of the panel-coach relationship (number 2, above), and the richness of the data coded to that node, this article examines that particular relationship exclusively. Please see the Appendix Table A3 for a partial coding schema (those nodes relevant to this study, excerpted from the macro coding tree), including the subnode detail under “panel and coaches” around which this article is organized.

For an audit trail from the research questions through data sources and analysis, please see the Appendix Table A4. For additional information regarding the original study’s design and methods, including

interview protocols and the full coding schema, please see Goldstein (forthcoming).

FINDINGS

This section of the article first presents descriptive detail of the workings of PAR and the PAR panel in Rosemont and then examines how the panel both supported the coaches to do their jobs and held them accountable. The section concludes with data demonstrating the shortcomings of the PAR panel structure in the face of enduring norms against accountability in education.

The excerpts from PAR panel meetings presented throughout this section are not intended as a model or best case; on the contrary, in places it would be easy to lodge critique about basic aspects of the discourse, such as expertise in coaching or meeting facilitation. The data presented here speak to the research questions—what did it look like, and how (if at all) did the presence of the panel affect the teacher evaluation process and outcomes? The argument is not that what occurred was ideal but rather that it was a significant departure from standard operating procedure.

PEER ASSISTANCE AND REVIEW AND THE PAR PANEL DESIGN

The first step in understanding PAR is understanding the role of PAR coaches. PAR coaches are not based at one school but rather are matched by school level and/or subject area to teachers across multiple schools. In this way, one master high school English teacher might be supporting and evaluating novice English teachers across four different high schools. Coaches typically visited each of the approximately 10 teachers in their caseload once every 1–2 weeks to conduct formal and informal observations, announced and unannounced, and to meet for planning, debriefing about observations, and a myriad of other support activities (see Goldstein, 2007b, for more detail). In addition to conducting formative assessments of coaches as part of mentoring activity, however, coaches were also charged with conducting summative personnel evaluations. PAR alters the traditional structure of teacher evaluation by not only involving teachers in the formal personnel evaluation of other teachers but also by often investing them with primary responsibility for the employment recommendation in place of principals.

Although teachers evaluating other teachers through the PAR coach role is novel to many educators and may be considered the defining aspect of PAR programs, one of the most important components of PAR

is the oversight panel. These panels have different names in different school districts but maintain the same basic structure.⁶ The districtwide joint teacher/administrator panel is typically co-chaired by the teacher union president and the director of human resources (or some other high-ranking district office administrator) and consists of a majority of teacher members. In California, for example, the state legislation specified a panel of five teachers and four administrators, with the teacher union president considered a teacher (Villaraigosa, Strom-Martin, & Alquist, 1999). The panel holds hearings several times a year, at which coaches provide reports about PT progress. At the spring hearing, and sometimes sooner, the coaches make recommendations about the continued employment of each PT, and the panel, when necessary, challenges the coaches on the evidence that they have provided to support their recommendations. A PT must meet specified quality standards within a set period of time, usually one year, or face removal from the classroom. This is determined by the panel based on the recommendation of the PAR coach, sometimes together with the principal, at the panel hearing. The panel's employment recommendation is then passed to the superintendent, who makes a recommendation to the school board, the ultimate arbiter of personnel decisions. In most PAR programs, new teachers spend one year in PAR and then revert to evaluation by the principal. For intervention cases, once a teacher has successfully exited the program, he or she returns to traditional evaluation by the principal.

In Rosemont, the PAR panel convened in the spring semester prior to the start of the PAR program to begin the process of hiring coaches and conducting program development. Once the program began in the fall, the panel continued to meet approximately once each month for a couple of hours for a business meeting and met four times during the year for panel hearings.

The panel hearings were the most significant and formal public examination of PTs' practice, and therefore also of coach practice, of the PAR program. Hearings ran all day for two days in November, January, and April—8:30 a.m. to 6 p.m. with a working lunch—plus a final day in early May for any appeals by participating teachers. The two January dates were actually separated by 3 weeks, with the latter day serving to revisit challenging cases presented earlier in the month. The atmosphere was formal and somewhat tense; a coach sat at one end of a conference table to make his or her report to the assembled panel members and other high-ranking (though nonvoting) district administrators, such as the superintendent, who were present at times. Coaches reported on PTs'

growth and/or problematic practice to the nine cross-district panel members, first with extensive documentation and then with oral presentations. In addition, the teachers' principal was typically present and would follow the coach's report with his or her own impressions, particularly regarding out-of-classroom matters.

The composition of the panel was hailed by most as an asset of the program. It included a cross-section from across the district—classroom teachers, teachers on special assignment, building administrators, and district administrators. In addition to the teacher union president, the teachers on the panel included an elementary bilingual specialist, a middle school language arts teacher, a high school English teacher, and a math coach. The administrators on the panel included an elementary principal, a middle school principal, the associate superintendent for the high school division, and the head of human resources. A number of people commented that if an observer walked into the room during panel hearings, he or she would not be able to tell which members of the panel were teachers and which were administrators. Noted one coach, "I think the panel members are fantastic. I really appreciate them. I really appreciate that there are teachers there, there are principals there, there are district administrators there and we're all having the same conversations, we're on the same level with conversations."

The coaches' work with PTs, including assessment documentation and presentations at panel hearings, was grounded in a modified version of the California Standards for the Teaching Profession (California Commission on Teacher Credentialing, 1997). Coaches were not experts in performance standards for teaching at the time they were hired. Prior to the beginning of the school year, the panel and coaches took a retreat devoted to professional development on coaching methods—which included the use of teaching standards. The coaches then poured many professional development hours during the year into becoming experts on the standards and into becoming calibrated among themselves in their use of the standards for evaluating to a rubric.

The coaches generated standards-based forms for their work with PTs. These included an Individual Learning Plan (ILP) that specified the standards and precise elements on the standards that were the focus of their work with PTs, as determined by the coaches and PTs together or by the coach independently if necessary. Observations of PTs' teaching would then generate evidence aligned to the ILP, and coaches' summaries of their observations would chart PTs' progress (or lack thereof) along those standards. Documentation of coaches' work with PTs—summaries of observations, ILP forms, coaches' ratings of PTs, and

evidence such as PTs’ lesson plans—was gathered into portfolio-like binders. Each panel member received an updated binder prior to the hearings.

Because time was a factor in the panel hearing process, the panel members decided that 91 PTs was too many to discuss in the two days allotted for hearings. They made the decision early on to bring to the full panel hearings only those cases that were “red flags”—those PTs whose year-end outcome was uncertain after the initial presentations at the November panel hearing. The rest of the PTs were presented in subpanel meetings called “quads,” smaller groupings of two coaches (who were considered “critical friends”) and two panel members who met periodically.

Of 88 beginning teachers in PAR, 24 were brought to the panel hearing in April (see Table 1). Of these, eight were renewed, including one who would continue to receive extra support the following year, and one who would be placed in a different school because of an unworkable situation between the PT and the principal. Four PTs were midyear additions to PAR, would continue in the PAR program for another semester, were deemed to be performing satisfactorily, and would therefore be revisited in the future in quad meetings. Eleven beginning teachers were not offered the opportunity to renew their contract for teaching duties in the district, including one who was released effective immediately, and one who was specifically told in writing that he was welcome to come back to Rosemont if he earned a teaching credential. One additional PT was nonrenewed, to be enacted by the principal rather than the PAR program, because of a fluke in employment status. Finally, the three veteran teachers in the program were removed from teaching but their cases were resolved prior to the April hearing through early retirements or reassignment to out-of-classroom responsibilities.

Table 1. Summary of Year 1 Participating Teacher Outcomes

Outcome	Beginning	Teachers	Veteran	Teachers	Total
	Renew	Nonrenew		Nonrenew	
Resolved in quad meeting	64				64
Resolved at panel hearing	8	12 ^a	0		20
Continue in program another semester (move to quad)	4				4
Resolved outside of panel (reassigned, retired)				3	3
Total	76	12	0	3	91

^a See Table 2 for more detail.

SUPPORTING THE PAR COACHES

The PAR structure enabled the coaches to be supported by the panel in their work. In a prime example of what Elmore (2003) calls “reciprocal accountability,” the panel that holds the coaches accountable must also provide them with support to do their jobs. Whereas principals are mostly on their own to evaluate teachers, the coaches had a community of educators with whom to discuss their work and consider decisions. Some coaches found the performance aspects of panel hearings anxiety producing but nonetheless expressed gratitude for the role of the panel.

What follows is an excerpt from the January panel hearing, at a point at which the conversation was focused on one of the three intervention teachers in the program that year. The coach and principal were both present, making reports to the panel. Some other coaches were also in the room. The PT was a high school science teacher. The coach described the teacher’s difficulty creating and maintaining an effective learning environment, particularly with directed teaching. The coach linked the need for whole-class instruction to the learning goals in the PT’s subject area and outlined the low level of learning taking place for students as a result of her inability to command a class’s attention. He explained that despite the teacher’s efforts, he was not seeing growth in her performance. At the point at which we enter the conversation, the coach is expressing the need to try something new.

Coach 1: I’ve sort of hit the wall. I need to go into a whole new approach. I’m feeling like I need to do a series of demonstration lessons. We’ve kicked around the idea of sending her someplace to observe, and there just isn’t anywhere. I know some dynamic teachers—but putting her into a place with dynamic teachers I don’t think will be useful unless she can be privy to what’s behind it. I’ll need to be there anyway, so I might as well do the teaching. So I’m thinking I’d do a series of constructed lessons where I share with her why I’m doing the planning I’m doing, and she sees it unfold. But the revelation that kind of hit me finally is that she needs to be treated as a first year teacher—the fundamentals aren’t there.

The panel members in turn ask questions of the coach and the principal. Panel Administrator 1 was the co-chair of the panel with the teacher union president.

Panel Administrator 7: I'm wondering about two things. First, what do the students say to you? Second, do you feel that there are some students that are being responsive to you being there?

Coach 1: I've tried to be basically a fly on the wall, so I didn't change behavior. Nevertheless, I have heard some feedback from some students, some of it is actually fairly adult thinking. "Are you here to help?" [laughter] I was thinking maybe I should create a venue to hear [from them] more.

Panel Administrator 1 [to the principal]: Do you think [the PT] would benefit from [the coach's] suggestion of demonstration lessons?

Principal: I think she's very receptive, she never says "oh that won't work." But the follow-through isn't there.

Panel Administrator 1: Maybe no matter what we do, change just won't happen, but we want to try everything we can.

Principal: I think it would be very helpful.

Panel members then chew on the idea of demonstration lessons, voice their concerns, and make suggestions.

Panel Administrator 7: Maybe you should teach and have her not be there, so she's not influencing the kids. You could videotape, so she could see how her students interact with someone else.

Panel Administrator 1: I think that's a good suggestion, because if she is there, the kids may turn around and go, "see?" They may embarrass her which might be counterproductive.

Coach 1: She's relatively callous to that.

Panel Teacher 5: Well she spent all that time as a paraprofessional, she was watching someone who could control a class. But clearly she didn't internalize it. I'm not sure where we go with that. I think it's a good idea, but is it going to work? Maybe we're just repeating. . .

Panel Teacher 6 [to coach]: But I heard you say you'd spend time with her around your thinking, all the preplanning.

Coach 1: I think that's the difference.

Panel Teacher 4: How would you translate the prethinking, the attitude, everything that will go into you developing this lesson? I think it's a good step, of reminding you of all the pre- and during and post-thinking that has to go on. But I'm not sure that translates to people just by watching. I'm a little bit worried about the modeling piece. Across the board.

Union President: What I've heard from [Coach 1] is that maybe we need to go back to ground zero. That there are some fundamentals here that haven't gotten through. This is very basic stuff that she would never have seen as a para, would never have had anyone draw her attention to it. Maybe she didn't get it in teacher training. . . . What I'm hearing is that there's a glimmer of progress, but it's really dim right now, and we'll be hearing about her next time.

As the conversation progresses, coaches in the audience also join in.

Union President [to coach in audience]: You had a question?

Coach 10: As a coach, I have a concern about going in and modeling a lesson, and then the next day the teacher comes back, and I don't know that kids will appreciate that.

Union President: [Coach 7]?

Coach 7: I'd be curious to see not only you talking with her about your planning but what she sees. Have her take notes. What does she see in the lesson you do? Internalizing it is yet another step, but just her being able to see it.

Union President: That's another reason to tape, so if she comes away with nothing, you can go back and revisit. What we're looking for is capacity for growth.

Finally, a panel member reminds the coach, and the group, of the importance of documentation.

Panel Administrator 2: I just want to ask about the ILP [the PT's individual learning plan]. These are evolving documents. Any time you think it's appropriate to modify, update the ILP, it's important to do so. We want the ILP to be a pathway for her to improve, but it's also due process and documentation for the program. The presence or absence of ratings that can be backed up with evidence are really important.

Ultimately, growth was not seen with this teacher, and the panel voted to dismiss her at the second January hearing. The teacher in turn chose to retire, leaving at the end of the year.

Dialogue at the panel hearings also served as a forum for identifying “red-flag situations” across the district. For example, the panel identified some cases of principals failing to give beginning teachers a sufficient opportunity to succeed, such as an assignment of four preparation periods across three classrooms on two different floors of a building. Extremely challenging situations like this complicated the coach's job of diagnosing and assessing a PT's practice and potential. In such cases, the associate superintendent on the panel addressed the situation with the principal directly and sometimes required that the conditions for the new teacher be altered.

Coaches were also supported in quads, the subpanel groups of two coaches and two panel members that met periodically. Coaches could look to quad meetings for guidance with programmatic issues, such as negotiating relationships with principals. In addition, coaches brought their most difficult PT cases to quad meetings to brainstorm and trouble shoot. One coach noted,

What the quad does is that at least you know when you are given such a short time to present the work that you've done with the PT [at the panel hearing], you know that two members of that panel have been privy to more conversation and have been able to go more in depth with the case. I think the panel can kind of look to those two members and think, well, they know a little bit more about this story and what you've tried. They would have more insight to how the case developed.

Coaches occasionally brought their critical friend—the other coach in their quad—to jointly observe a PT's classroom for additional suggestions or assessment.⁷

Across the board, coaches would have liked support from the panel to be improved during PAR's initial implementation. Most complaints

involved the ambiguity that surrounds policy implementation (Baier, March, & Saetren, 1988): Coaches wanted more directions, clearer directions, and more timely directions. One commented, “It would have helped if they could have said to us do A, B, and C this way.” Another said, “Just tell me, you know, I just want you to tell me this is the way it is.” A teacher on the panel concurred:

I think the program had to start under so many compromises, but the coaches did an admirable job in a difficult set of circumstances. I don’t think they were given as much support as they needed, and I think we went into it too quickly. I don’t think that our understanding of what it would take to help someone really know what is the difference between mentoring and coaching, and how that looks [was sufficiently developed].

Because many of the programmatic details were still being worked out, one of the questions that the panel asked coaches during the coach selection process was whether they would be comfortable operating with a lot of ambiguity. Despite this warning, coaches were often frustrated and wanted more support—often defined as clarity—in the first year of the program.

HOLDING THE PAR COACHES ACCOUNTABLE

Perhaps the most profound structural distinction between regular teacher evaluation and PAR is that the coaches must defend their employment recommendations to colleagues. They must come before a group of other educators and say, “I believe we should rehire this teacher and here’s evidence why,” or alternately, “I believe we should *not* rehire this teacher, and here’s evidence why.” In addition, they have to demonstrate that the PT received sufficient support and had an opportunity to improve. The teacher union president identified the role the panel played in holding the coaches accountable at the hearings:

The coach has to make the recommendation. The panel has the responsibility of making sure that recommendation is supportable. That’s what the questioning is about, that’s what the documentation is about. If a coach is making a recommendation, they have to be able to justify it to the panel, and if they can’t justify it to the panel, then it gets overturned or they get sent back [to gather more evidence].

Noted one coach, “I’ve been really impressed with the panel. I think the panel has been very clear on what they see as evidence, what they find are good individual learning plans [for PTs] and what they think is growth and not growth.” One coach commented that having to come before the panel with evidence to support a recommendation to either retain or fire a PT “keeps you honest” regarding what support has actually been provided to the PT and what evidence of growth has actually occurred. The comments in the following panel hearing excerpt, directed at the coach, are representative:

Panel Administrator 2: It sounds like [the PT] has lots of content knowledge, can bring a lot of activities to class, but it’s not tied to goals and assessment.

Panel Teacher 4: Except that I’m going to ask, did you provide support in those areas?

The panel that holds the coaches accountable for teachers’ evaluations must also provide them with support. Similarly, part of holding the coaches accountable is ensuring that the coaches have provided adequate support to the PTs, who are in turn being held accountable for classroom teaching. It is worth noting that a companion study (Goldstein, 2007a) found PTs largely positive about the program.⁸

Table 2. Summary of Nonrenewed Beginning Teachers

Coach’s original written recommendation was to nonrenew	Coach recommended nonrenewal after challenge by panel	Revert to Principal	Total
7	4	1	12 ^a

^a One beginning teacher reverted to principal recommendation because of a technicality in employment status, therefore 11 beginning teachers were officially nonrenewed by panel recommendation.

Of the 24 cases brought before the panel in April, four cases across two coaches involved the panel seeking more evidence or otherwise challenging the coach’s initial recommendation (see Table 2). Although the panel ultimately supported the coaches’ recommendations in all cases, a close analysis of the panel hearing transcripts shows that in these four cases, the coaches’ recommendation actually changed as a result of the panel hearing. In these cases, the coaches recommended renewal in the written assessment documentation but gave a wealth of negative evidence orally at the panel hearing. In one of the cases involving an uncredentialed teacher, after both the coach and principal had discussed the PT’s performance, the following exchange occurred:

Panel Teacher 6 [to coach]: Why are you recommending that he be “renewed”?

Coach 10: I put down renew with proof of enrollment in a credential program.

Panel Teacher 6: It sounds like if he were in school while trying to teach, *nothing* would get done.

Principal: That’s the impression I get.

Coach 10: Well, if he were in a program, that would at least show us that he’s interested in being a teacher, and maybe once he were taking classes, lights would go on about what teaching entails.

Panel Teacher 4: I don’t think at this point it’s up to us to mother [the PT]. I think we make a recommendation based on what you’ve observed. If [the PT] has not met the requirement for rehire, then he should not be rehired.

The coach noted in an interview a few weeks later, “I was tap dancing around giving a decision of nonrenewal, and they asked me directly, what is the evidence for keeping this person? And I really didn’t have enough. They held me accountable, and that was appropriate.”

Similarly, in another of the four cases, the panel identified the inconsistency between the coach’s recommendation of continued employment and her verbal presentation, which revealed many concerns about the PT’s performance:

Coach 4: This teacher I’ve gone back and forth on, but I did recommend him for rehire based on his strength in science. He’s not a very good planner or record keeper, but he does love science. I asked him for his objectives or purpose, and he looked at me like I’m asking some weird question.

Panel Administrator 2: If he doesn’t have objectives, how does he know what to teach?

[laughter]

Union President: I’d like to see the summary evaluation reflect

more of what you've just said about where he is. Everything you've written here supports the notion of reelection. In fact where you have placed him indicates that he has met standards, and you're now saying that he hasn't.

Coaches commented on the difficulty of putting a negative assessment in writing or of having the weight of the decision fall on them. The following exchange occurred in a third case, after both the coach and principal described the negative performance of a PT to the panel:

Union President [to coach, with incredulity]: Your recommendation [on the documents] was for renewal...?

Coach 10: Contingent upon [the principal's] assessment of [the PT's] professional accountability and reliability.

Union President: There's a disconnect between the numbers you have put on this page, and the recommendation.

Coach 10: Well, we coaches have a real hard time with putting on this paper, "do not rehire" . . . we do.

Union President: But that's what you're supposed to do.

Coach 10: I know, I know . . . I would support not rehiring him.

Union President: When I was listening to you, I was listening to hear something that would justify keeping him.

Panel Administrator 1: Me too.

Union President: Where I thought you were going to go is his commitment, that he had signed up for a credential program. But it doesn't seem to make any difference.

Coach 10: I would say not renew him. There's a big learning that I hope will take place, he is taking classes, but I don't know that he will be any more effective than [he is now based on] what he said to [the principal], that he wants students who are self-motivated, where he just writes problems on the board.

Union President: I hear a recommendation of not reelect. Is there a motion?

Panel Administrator 1: I'll move that . . . I would hope that this page can be revised to show that the coach supports the panel's recommendation not to rehire.

After completing the voting process, more discussion ensued.

Panel Teacher 3:[The coach] raised a really important issue, that coaches are struggling with needing to put down whether to rehire or not.

Panel Teacher 5 [to the coach]: You shouldn't bear the burden of that at all. The weight of that shouldn't fall on you.

Coach 10: That's the discussion we had [amongst ourselves] as coaches, whether we were just presenting evidence and you decide, or whether we had to make a recommendation. That's why mine say things like "contingent."

Panel Administrator 9 [to coach]: I think you bring up a good point for us to discuss tomorrow. How can we support you better to do this. Let's face it, principals have the same problem.

The coaches therefore gave voice to what researchers have long observed in principals (Bridges, 1986, 1992). A disparity is typically seen between formal and informal teacher assessments. In a study of principals across Virginia, for example, principals formally identified 1.53%–2.65% of teachers as "incompetent" but informally identified 5% as "incompetent" (and only 0.1% were actually dismissed; Tucker, 1997). The dismissal rate in Rosemont was also 0.1% (3 out of 3,000) in the year prior to PAR implementation. Conducting performance evaluations, whether in teaching or other occupations, is often a difficult process. Evaluators often doubt themselves or seek to avoid confrontation, thereby failing to give either feedback that would actually help improve performance, or a negative summative assessment despite cause for one (see for example Bridges, 1986; Danielson & McGreal, 2000).

The backdrop to these excerpts is a shift from isolated decision making to professional community. In dialogue with other educators, with the panel as a mirror to their assessments, and with documented evidence of teaching performance over time, the coaches altered their official

recommendations to match their informal assessments—and in several cases, they appeared relieved to have the panel point out the obvious. In this way, the PAR panel “closes” the system (Halverson, 2003) by creating formal lines of communication and reporting between coaches and panel members regarding the current teaching quality and potential future promise of PTs. In doing so, the transparency of the process was increased, and some teachers who would have otherwise fallen through the cracks—just as they often do with traditional teacher evaluation—were removed from teaching.

FALLING THROUGH THE CRACKS, STILL

Although the panel heard only 24 of the 91 PT cases at the January and April hearings, most of those involved still felt that the hearings were rushed, not allowing sufficient time to go into the depth they would have liked. Some coaches, although sympathetic to the amount of work involved, wanted panel members to be more informed. One commented, “[Panel hearings] felt generally okay but rushed. I would like it if in the future they could read the documentation a little more carefully. I know it’s a lot, but it would help them listen to the stories a little better and ask better questions of us.” Some coaches therefore reported feeling that the panel was a rubber stamp on their decision about a PT. One noted,

In every case pretty much it was what the coach recommended, and maybe it should have been that. But when I think about the questions they asked about [a PT whom I recommended for non-renewal], they didn’t really hold my feet to the fire. They just didn’t really go after me in ways that they could have. I would have probably just said the same thing and it would have had the same result, but there would have been more a feeling of somebody really advocating his side.

A second coach echoed the sentiment:

In fact, I think the only, not even criticism, but suggestion I have to the panel is they need to be a little tougher I think on us, especially when it comes to really nailing people down at the end. I felt it was kind of a rubber stamp situation. I mean I can understand why it would be because they can only go on the data we present. But I really think when it comes down to telling somebody, “No, we’re not going to let you return,” I think they need to have read chapter and verse and really thought about it.

Although the data reveal movement in a direction toward increased transparency and accountability, there was certainly room for growth.

Time was only part of the issue. Although the existence of the panel increased transparency, much may still have been obscured, with potentially deleterious effects on accountability. One particularly poignant case emerged of a red-flagged PT who was renewed but perhaps should not have been. It involved a coach and a principal who were both highly respected by the panel, which may have contributed to the panel's easy agreement with the recommendation to renew. From the hearing:

Coach 2: In standard 2, she is getting closer to meeting standard. . . . The main issue is standard 2, an environment of fairness and respect. Her students do not use respect. They're just not very respectful to each other, to her. Her classroom is not one you walk into where you say this is a good climate. It's borderline. It was a bad climate, they were out of their seats, throwing papers, those things are not happening anymore. It's the bare minimum.

Panel Teacher 5: Does she like the children?

Coach 2 [opens mouth]:

Principal: No. [*laughter*] I'm sorry [Coach 2], she doesn't. She's angry at them.

Coach 2: She has resentment, yes. She resents. . .

Principal: . . . that they're naughty.

Coach 2: She resents them, but she tries. She works very hard—she has gone to every [professional development session]. She always meets with me, takes on my suggestions, is very concerned about her professional growth and development. . . . If you look at her lesson plans, they are typed with every single step written out.

Union President: Let's hear from [the principal].

Principal: Every bone in my body says she doesn't have the right stuff. However, it's her first year, she works really hard. There is something to the fact that the chemistry of this class is really bad. I feel that she deserves another shot at it with support. . .

Union President: So your recommendation is for retention?

Coach 2: If there are ways of supporting her next fall.

A discussion ensued about the support that could be provided to this PT for the following year. BTSA refers to California's statewide mentoring program (Beginning Teacher Support and Assessment), which in Rosemont provided far less intensive support than that provided by the PAR program. Continuing:

Union President: She is a BTSA candidate.

Coach 2: Having been a BTSA coach [myself], she needs more support than that.

Principal: BTSA is too advanced for her.

Coach 2: I want to make that clear. I did not write BTSA down because BTSA is not enough.

Union President: Well, I'm having a little trouble with that. Because if the recommendation is for that much support, what is the recommendation, really?

Panel Administrator 2: Intensive support.

Union President: But we know that that doesn't exist [in the district].

Panel Administrator 2: Yet.

Union President: But if we're saying she can't function without intensive support, then we're saying she can't function.

Panel Teacher 4: No, where she's scored on the documents, she's meeting standards.

Union President: She's not meeting classroom management standards, and what she is doing, she's doing with intensive hand holding.

Coach 2: If she's not going to get the type of support I'm saying

she needs, then I need to come back tomorrow or a different time when there's more time, because this is a case where I'm not sure.

Ultimately, the recommendation was to rehire the PT with continued support, "as yet undefined." The PT was in fact assigned a full-time mentor the following year, a rare occurrence, although her performance problems continued. During an interview a few weeks after the panel hearing, the coach reflected on the case:

I'm still wondering if we did the right thing . . . I did present evidence that didn't sound too good and [the panel members] didn't ask for more. They didn't have questions about it. There wasn't communicated a strong concern or a strong, hmmm . . . I thought if there were more questioning it would draw out what the *final* final [recommendation] would be. I came in with a recommendation and I know for some coaches that recommendation changed. It kind of sounded like because [the principal] and I were like, we both see it this way, we both think these are her strengths, we both think these are her weaknesses, they were like okay, fine, all right, time's up.

The tension between the time needed to fully question practice and "the press simply to get on with it" (Little, 2003, p. 940) is a common facet of professional communities of practice. This coach identified the dynamic between the coach and the principal as a key factor in the outcome. In this case, she and the principal had a close working relationship, and although both believed that the PT did not "have what it takes" to be a teacher, they both reported wanting to err on the side of "compassion." The coach speculated on how the outcome might have been different with a different principal, specifically the principal for other teachers in the coach's caseload:

I tell you one thing, if [this PT] were at [a different school] with [a different principal], she wouldn't have made it through. That's how the collaboration kind of infests itself. [The other principal] would have been like, "[coach], she's not coming back here," and I would have been affected by that. Where [the PT's actual principal] and I worked on the other end, our decision was rooted mostly in having compassion for this person and giving her more time and not going with our gut instinct of this person isn't effective.

Finally, the coach reflected on her own process and development in her new role, and the implications for accountability:

I needed support in the opinion. It's like somebody needed to shake me up a little bit to get to [a nonrenewal]. Now what I need to do next year is I need to bring a lot more coaches into those classrooms or I need to bring another coach consistently into that classroom, or I need to get the principal in the classroom a lot more. Why, why do I need that? I don't know. I think I just do right now. But maybe next year I won't because I've had a learning experience. . . . It would be interesting to look at the statistics of how many first-year teachers are let go by administrators, and compare that to what we did and see if there's a difference, because I don't know if there's a difference or not. Had I let [this PT] go, I think there might have been a difference. Had I chose to let her go.

The coach highlighted here was unusually perceptive and candid regarding her own hesitance to give a nonrenewal recommendation. At least half of the coaches, however, commented on the significant role the panel played when it came to getting them to recommend dismissal. Coaches believed that the panel needed more time for hearings, but other factors were certainly their own reluctance to recommend dismissal and their desire for the panel to remove some of this burden, as historic norms in education precluding teachers from passing judgment on other teachers played out (Little, 1982, 1988; see also Goldstein, 2004). In addition, the coach just quoted argues here that principals' habituated behavior regarding teacher evaluation has an effect, to the extent that principals and coaches work together on the task.

Although PAR involved a much larger number of people in a more transparent teacher evaluation process and broke down some of the isolation around both teachers' practice and evaluators' decisions, coaches clearly felt that even more transparency was needed. PAR is no panacea, and the challenges to holding teachers accountable for practice remain daunting.

DISCUSSION

The article began by asserting that a primary goal of instructional leaders is to design organizational structures that promote high-quality teaching. Typically, school district structures occur in one of three ways: by default, by coincidence, or by inheritance (Halverson, 2003). Actively

designed new structures are rare, and designing a new structure for teacher evaluation is to interrupt a century-old and taken-for-granted institutionalized facet of school district life (Scott, 1995). PAR would be interesting to explore if for no other reason than its existence as a case of organizational change, and the article has given a picture of what an oversight panel for teacher evaluation looks like in one instance of practice. The data, however, suggest that a fairly radical shift in teacher evaluation outcomes and process occurred.

As one key “artifact” (Halverson & Zoltners, 2001), or structure, in a larger system of practice (Halverson, 2003) for teacher evaluation with PAR, how did the PAR oversight panel make such a difference? This section of the article entertains a few possible answers. It first examines how the structure of the PAR oversight panel, as seen through the examples presented in the data, generated transparency in the evaluation process. Second, the section explores how this transparency enabled collective responsibility for professional standards, and hence increased accountability for personnel outcomes.

DESIGNING TRANSPARENCY

The PAR evaluation process in Rosemont, and in particular, the presence of an oversight panel, appears to have increased transparency in two central and related ways. First, panel members were involved in teacher evaluation in addition to coaches and principals, creating a professional community of educators focused on the examination of practice. A communal teacher evaluation process involving dialogue increased transparency because it forced what would otherwise be an isolated evaluator to explain his or her assessments to others—thereby bringing the evaluation into the light of day. Second, there was a paper trail to document the assessment work that had occurred, creating data, providing the focus for professional conversations, and codifying practice. Documenting the assessment of teachers increased transparency because it allowed the panel’s conversations to focus on teachers’ practice—on observed behaviors rather than opaque evaluator opinion.

Creating Professional Community

Examining the Rosemont phenomenon from a professional community perspective emphasizes that PAR broke down prior levels of isolation. The need to reduce teachers’ isolation and open the doors of classroom practice has become axiomatic (i.e., Feiman-Nemser & Floden, 1986; Little, 1982; Miles & Darling-Hammond, 1998; Rosenholtz, 1989), often

referred to as a need to alter the “egg crate” model of schooling that isolates teachers in separate classrooms like eggs in a carton. Although PAR opened the doors of teachers’ classroom practice, what the panel data show is that the doors to support provider and evaluator practice were opened as well. The work of teaching and the work of evaluating teaching were both opened for examination by the panel. For example, the conversation about the coach considering model lessons for the veteran intervention case shone a light not only on the classroom teacher’s practice but the coach’s practice as well. The panel members considered whether modeling was likely to be a successful strategy for working with the teacher and, if so, how it might best be done. At the center of the conversation were questions about the content of good teaching practice—effective planning and execution of whole-group instruction—and how to break that down and teach it to someone whose practice is below standards in that area.

The panel hearings became a site for professional learning community, for a group of educators coming together around questions of practice. Little (2003) identified the markers of “teacher learning communities:”

the groups demonstrably reserve time to identify and examine problems of practice; they elaborate those problems in ways that open up new considerations and possibilities; they readily disclose their uncertainties and dilemmas and invite comment and advice from others; and artifacts of classroom practice (student work, lesson plans, and the like) are made accessible. In all these ways, the groups display dispositions, norms, and habits conducive to teacher learning and the improvement of teaching practice. (p. 938)

As the PAR panel included administrators as well as teachers, it could perhaps be considered an “educators’ learning community” instead of the more standard “teacher learning community,” but the similarities are evident. The panel had set, reserved time to meet and discuss problems of practice. They discussed the problems of practice in ways—however new and uncertain—that involved disclosing uncertainties and dilemmas (“*we coaches have a real hard time with putting on this paper, ‘do not rehire’ . . . we do.*”). The coaches brought forth artifacts of PTs’ classroom practice and artifacts of their *own* practice (ILPs, observation summaries, and the like) around which to focus conversation and inquiry. In short, the panel members and coaches engaged in dialogue around the technical core of their work—the practice of teaching, coaching, and assessing. This engagement was indeed somewhat messy, as those involved worked to

figure out the nature and boundaries of their task. In doing so, however, they may have avoided (perhaps unintentionally) what Seashore Louis (2006) warned is a tendency to develop professional learning communities “like a new reading curriculum”—as superficially adopted ideas with short shelf lives rather than as internally developed programs with depth and commitment. A teacher on the panel gave this example:

I think by the end we nailed it down. There was so much ambiguity [at first]. I think by the end when you look at reports like [Coach 2’s], where she bullets, “here’s the standard, here’s the three things that I’m going to look for that would be a demonstration that you’ve mastered the standard, and then here’s what’s happening at each of those.” Very clear. But we didn’t get to that point until after we’d gone through some serious adjusting. But I think we could have. Oh, maybe we couldn’t have. Maybe that’s just the whole point that you have to sort of go through it before you know.

Although panel members and coaches complained about lack of clarity and ambiguity along the way, they were forced to face core questions of practice: What do we consider good teaching? (What does it look like in a novice versus a veteran? In a credentialed versus an uncredentialed novice?) In addition, what do we consider good coaching, and good assessment of teaching? How do we generate data about observed teaching practice so that a group of educators can discuss it and make appropriate decisions about it? For many of the educators involved in PAR in Rosemont, this was the first time they were able to engage in such conversations, and certainly it was the first time that they engaged in them as a group of administrators and teachers from across the district, as colleagues. One teacher on the panel noted, “As a new panel member, I have enjoyed being a part of this new adventure. It has put me, along with this team of panel members, in a learner mode. It’s been very exciting.”

It seems likely that the dialogue undertaken by the professional community generated “better” (i.e., more thoughtful) decisions about teaching practice and employment. A key difference between this particular professional community and the teacher learning communities that constitute the literature is the formal authority possessed by the group. It is a distinction taken up in the following discussion of accountability.

Codifying the Knowledge of Practice

A codified, shared, and “specialized” (Freidson, 1994) or “abstract”

(Abbott, 1988) knowledge base defines and demarcates professions from other occupations. Teaching has historically been considered a “semi-profession” (Etzioni, 1969) without this knowledge base, and Elmore (2006) identified the lack of codified knowledge of practice as the biggest problem currently facing education—in part because it leaves teachers without power as experts in the field (also Lortie, 1975). Part of the historic problem with teacher evaluation, in turn, is that without clearly defined standards for good teaching practice, there is subsequently no clear agreement among educators about how teaching should be evaluated. Neither PAR nor an oversight panel inherently codifies teaching practice, and the program could exist without standards-based evaluation.⁹ Yet the work of PAR in Rosemont was intimately tied to standards for good teaching and to conversations about the meaning of the standards.

The pressure of the panel hearings, and specifically the panel’s expectations that the coaches’ assessments would be standards based, simply created more data and documentation. This stood in contrast to the ubiquitous “I know good teaching when I see it” that has plagued much of traditional teacher evaluation. Strong evaluation systems include established standards for performance, rubrics, and evaluator training for interrater reliability (Tucker, 1997), yet many teacher evaluation systems nationally lack these components (Loup et al., 1996). Noted one teacher on the panel,

Teachers and those of us in the teaching profession need to figure out what teaching standards are so we can articulate them to the public and to the universities who are training teachers. We can’t just say we want good teachers. What does that mean? That’s one thing I love about the standards in any area, teaching and content areas. They require talking to other people. There is no way you can say, okay, here’s a standard, blah, and then everybody just follows it. Someone will always say, “well, what does that mean, and how would you know it when you see it, and what would I be doing, and what would you be hearing, and what would the students be doing, and what would a parent who came in the room say?” It requires that communities talk together.

Standards created the focus for the professional conversations about practice that over time created professional community.

In addition, teaching standards or “protocols of practice” (Elmore, 2006) depersonalized the process, creating an evaluation that focused on the practice rather than the person. The teacher union president noted,

We're trying to institute standards for teaching, so that people will be playing on a common playing field, with common rules. Hiring and firing decisions would be made centrally. They would be based upon standards rather than the whim of a particular individual. Who has defined what good teaching would be [until now]? Who has defined what the standards would be? Who has defined that has been the individual principal at the 100 sites. So, we have had kind of like Italy before Garibaldi. We have 100 separate standards. None of them written, of course.

The superintendent emphasized the importance of agreed-on standards:

Can it really be that [a coach and a principal] are on opposite ends of the spectrum of whether a teacher should be retained or not? I've got to believe that as professionals we are going to have the same criteria, we're going to be using the same observation forms. We're going to be looking for the same end results which is student learning and how that interaction with that teacher is occurring in those classrooms. And so the data is going to be the basis on which the decision is made. It's not just what the coach says versus what I say as a principal. It's based on evidence. Somebody is being very subjective if they can't look at the evidence and come to the same conclusions.

Coaches tended to speak with more authority than principals at panel hearings precisely because of their greater facility with standards language (see Goldstein, 2004). Their recommendations carried weight to the degree that they were transparent—clearly documented, supported by evidence, and grounded in practice rather than personality.

A Caveat on Ambiguity

It should be noted that the PAR evaluation process was certainly not entirely transparent. Ultimate responsibility for personnel decisions was actually somewhat ambiguous. The decision process was spelled out: PAR coaches make a recommendation to the panel, who pass their recommendation to the superintendent, who passes her recommendation to the school board. The school board has final say. Yet within this protocol, many practitioners felt awash in ambiguity; coaches and even some panel members complained that responsibility for "the final say" was unclear. I have argued elsewhere (Goldstein, 2004) that this ambiguity was a

weakness of the program, which may seem to contradict the current argument that the process increased transparency. It therefore seems important to note the distinction between the decision-making process (how decisions happen) and responsibility and accountability for those decisions (March, 1994). In the Rosemont case, PAR increased the transparency of the decision-making process; the teacher evaluation process involved more eyes engaged in more scrutiny of more data. At the same time, PAR generated ambiguity around ultimate responsibility for evaluation decisions. March argued that decisions happen through complicated processes in which it is often not possible to specify where or how a decision occurred. Although individual ownership of accountability for teacher evaluation decisions may have been ambiguous, the argument taken up below is that the evaluation outcomes demonstrated that accountability nonetheless occurred, regardless of this ambiguity—and perhaps even partly because of it.

FOSTERING ACCOUNTABILITY

Despite the mentioned shortcomings, PAR in Rosemont created a process that increased the transparency of some critical components of teacher evaluation. That transparency was the necessary precondition for professional accountability. Through their dialogue and over the gathered documentation, the panel could hold the coaches and one another accountable for their assessments of PTs' classroom performance. Although one might argue that superintendents are similarly responsible for holding principals accountable for the support and evaluation of teachers, it is well documented that administrators typically do an ineffective job with these tasks. However, data from Rosemont's PAR program suggest that accountability actually occurred. The coaches had to demonstrate to the panel that participants had received sufficient support to make success a possibility. They then had to defend their recommendations to keep or dismiss a teacher by providing evidence demonstrating growth or lack thereof on performance standards. The resulting personnel decisions were a marked departure from the norm: 12.5% of beginning teachers (11/88) and 100% of veterans (3 out of 3) were no longer teaching in Rosemont after the first year of the program.¹⁰

After a century of a hierarchical educational system that places administrators above teachers in the chain of command, many people believe that teachers cannot be trusted to conduct teacher evaluations. The Rosemont case shows that, at least under certain conditions, this concern is unfounded. The argument here is that the oversight panel—and the transparency it generated—was a crucial piece of the PAR process and

outcomes. First, the presence of the panel turned personnel decisions into group rather than individual decisions. Second, these group decisions generated collective responsibility for professional standards.

Distributed Responsibility for Evaluation

The PAR panel created group personnel decisions based on the professional dialogue that occurred and the data produced by the coaches regarding PTs' classroom practice. A distributed perspective (Spillane, 2006; Spillane, Halverson, & Diamond, 2001) on the role of the PAR oversight panel for the process and outcomes of teacher evaluation suggests that a dispersion occurred that allowed a group rather than an individual to make personnel decisions. Responsibility for teacher evaluation was stretched over the coaches and the panel members. Coaches were responsible for the support and evaluation of classroom teachers, whereas the panel *held the coaches accountable* for that support and evaluation in a case of distributed accountability (Goldstein, 2007c). Although certainly the coaches' recommendations weighed most heavily, the panel structure created conditions in which a group of people, rather than an individual, ultimately participated in making employment decisions.

Involving more people in the teacher evaluation process through PAR appears to distribute both the responsibility for teacher evaluation and the onus of accountability. Individual coaches doubted themselves when it came to giving a negative evaluation, just as individual principals do. The group, however, served as a transparent "window" (Webster, 1983): assessments regarding observed practice shone through clearly, unobscured by that doubt. In the face of colleagues sharing in one's opinions of practice, the coaches became more able to make the tough decisions necessary for accountability. When it came time to make decisions, the group was more confident in their decision-making process precisely because there were more people involved, more eyes on the evidence of practice. The coaches were engaged in collective responsibility for professional standards, but in those cases (or in most of those cases) in which they seemed likely to fall short, the panel served to compensate.¹¹

In addition, the oversight panel dispersed the weight of personnel decisions; coaches could take refuge in the "group" decision, even as their recommendations were the central factor in those decisions. This perhaps contributed to coaches appearing "rougher" on PTs and the fact that there were more dismissals; the coaches did not have to function as individuals, but rather the panel backed them up, and they could go to PTs and say, "the panel decided this." Ultimately, consensus diffused responsibility for the harder decisions. One might view this negatively, as

everybody looking to someone else to take responsibility, and the buck stopping nowhere. The group as a whole, however, appeared to display greater responsibility, demonstrated by their willingness to act when historically there has been little.¹²

Collective Responsibility for Professional Standards

Collective responsibility for standards is a key element of both professionalism and professionalization (Darling-Hammond, 1990; Englund, 1996). To view the Rosemont PAR panel through a professions frame is to recognize that a fundamental shift occurred in both. The data demonstrate teachers' ability and willingness to hold colleagues to professional standards (professionalism), as well as a shift in the locus of authority for personnel decisions (professionalization).

Teachers' internal commitments to upholding standards—their willingness to make tough calls regarding employment decisions and their competence in doing so—is a crucial component of professionalism (Freidson, 1994). Some coaches were able to come before the panel and recommend dismissal of a PT. Others were hesitant despite the evidence they presented, as seen in four cases in particular as presented in the findings. In these cases, the panel questioned the coaches' practice, meaning that they pushed the coaches to reconsider their assessments. The coaches and the panel members were seen to be willing to hold PTs to established standards for teaching practice. The hearing transcripts reveal that teachers on the panel, including the teacher union president, were as vocal as administrators (if not in fact more so) regarding holding classroom teachers to high standards. The willingness of these educators to hold colleagues accountable was directly related to the transparency of the group process. The presence of more eyes and ears and opinions, and the presence of more data on which to base those opinions, created a more confident decision-making process on which a collective professionalism could rest.

Teachers' authority in matters of gatekeeping and quality control is a crucial component of professionalization (Van Maanen & Barley, 1984). PAR granted teachers such authority through both the coach and the panel roles. The teacher union president highlighted the significance of PAR:

If you want to become a beautician, you are entering a profession, and the profession of being a beautician is controlled by beauticians. The governor appoints beauticians to the commission that determines the standards for entry into the profession

of being a beautician. What a radical notion that members of the profession would define the criteria necessary to exhibit that profession. Education, though, is too important to be left to teachers. You have to be an administrator, or better yet, a politician, to define what good teaching is. There are lots of us who don't think that that's valid, and we're fighting it, and PAR is part of that fight.

Prior work (Goldstein, 2004) argued that the significant shift for teacher professionalization with PAR resulted from coaches making employment decisions in place of principals. Recall, however, that the PAR panel in Rosemont, and by state law across California, comprised five teachers (including the union president) and four administrators. An analysis of the panel's role in PAR evaluation is also ultimately a story of shifting authority relations that placed teachers on the panel in new positions of leadership. PAR in Rosemont was not teacher evaluation without hierarchy but began to replace the hierarchy of individual administrators with the hierarchy of a panel—a joint panel of teachers and administrators that gave teachers a majority voice and weighed the input of teacher leaders (the coaches) most heavily.

The teacher union president's role as co-chair of the panel had clear implications for professionalization. By state mandate, the PAR program was a joint project between the district and the teacher union. Lawyers hold collective responsibility for professional standards through the bar. Doctors hold collective responsibility for professional standards through a board. The professional association of teachers, their union, has not historically held any equivalent power. Whether teachers themselves traded this professional authority for the right to bargain as workers over wages, or whether this authority was taken out of their hands long before unionization by the forces of bureaucratization, is a matter of opinion and political inclination (see, e.g., Callahan, 1962; Lortie, 1975). An oversight panel for teacher evaluation co-led by the teacher union president, however, clearly signals a radical shift in the potential role of teacher unions in setting and maintaining standards for the profession (Kerchner & Cauffman, 1995; Kerchner, Koppich, & Weeres, 1997). Principals were sometimes surprised to find the teacher union president "on their side" at PAR panel hearings. Teachers opposed to PAR typically complained that the union should not be involved in "management" responsibilities. One way that the PAR panel fostered accountability, however, was through the central role of the teacher union. By putting the teacher union and district, or teachers and administrators, together in a professional community of educators focused on relatively objective

measures of the quality of teaching practice, the teacher union moved from defending individual teachers to defending the profession of teaching (Kerchner & Mitchell, 1988).

Elmore (2006) has argued that there is a profound misconception in education regarding the meaning of *professional*. Professional is often understood as autonomy, and anything that compromises autonomy—such as an oversight panel or proscribed standards for practice—is viewed as anti-professional. Yet professions are made up of people who have autonomy within proscribed protocols of practice, and in a profession, people who do not follow the protocols are excluded from practice. Professionals are empowered by virtue of the authority granted to them by society to specify these protocols of practice (also see Benveniste, 1987; Van Maanen & Barley, 1984). In other words, a key part of professionalism is oversight, in which structures are in place to guard and defend standards of appropriate practice. PAR and its element of the oversight panel is therefore a more professionally designed model of teacher evaluation—more professional than bureaucratic oversight of teachers by administrators and more professional than teachers practicing without oversight. This is due in part to the process, because teachers themselves are conducting the evaluations. It is also due, however, to the outcomes, because the model appears to more rigorously guard and defend standards of practice. In this way, collective responsibility for professional standards becomes the link between professionalism and professionalization, and a lever for both.

CONCLUSION

Creating and maintaining the conditions for high-quality teaching certainly involve far more than firing underperforming teachers. Providing teachers with sufficient support so that they have the opportunity to be successful is perhaps the most critical piece of the teacher quality challenge (Fideler & Haselkorn, 1999; see Darling-Hammond & Sykes, 1999), and the support provided to classroom teachers through PAR is discussed elsewhere (Goldstein, 2007b). Nonetheless, removing underperforming teachers has been a historic weakness in public education (Bridges, 1986; Tucker, 1997).

It is therefore notable that the teacher evaluation process appears to have been strengthened by the increased transparency and distribution of accountability seen in Rosemont. Mechanisms affecting the individual evaluator were modified by the presence of the group, in the form of the oversight panel. Isolated decision making was modified by the presence

of a professional learning community, and the relatively sparse data on which principals' evaluations are based were replaced with a collection of standards-based evidence gathered over time. As a result, the group was probably wiser than any one individual and was certainly more confident. The group proved to be more likely than an individual to act in the interest of upholding professional standards of practice, and did so by holding evaluators accountable for holding teachers accountable.

Some school districts may lack the contextual factors to make PAR possible. The lesson of transparency is nonetheless valuable for a more traditional teacher evaluation system. District leaders can consider structures like the PAR oversight panel to support principals' teacher evaluations and hold principals accountable for their evaluative decisions. At the same time, the article will have failed if at this point anyone is inclined to rush out and simply start creating oversight panels. I hope it is clear that the structure alone is likely to be insufficient and that meaningful implementation requires educators who are empowered as professionals. Professionals examine and support their colleagues' practice and are given the authority to take needed action when, in their judgment, that practice is outside appropriate standards.

Acknowledgements

This research was supported by grants from the City University of New York's Professional Staff Congress and Faculty Fellowship Publication Program (FFPP). Thank you to Richard Halverson, the members of my FFPP writing group, and two anonymous reviewers for their feedback on earlier versions of the manuscript.

APPENDIX: STUDY DESIGN AND RESEARCH METHODS

Table A1. Observations—Number of Days (total hours)

Data Source	Spring Year 0	Fall Year 1	Winter Year 1	Spring Year 1	Total
Panel meetings	3 (8.5)	3 (6)	1 (2)	2 (4)	9 (20.5)
Panel hearings		3 (28.5)	2 (19)	3 (21)	8 (68.5)
Coach meetings		17 (108)	7 (42)	8 (48)	32 (198)
Coach professional development		3 (24)			3 (24)
Total	3 (8.5)	26 (166.5)	10 (63)	13 (73)	52 (311)

Table A2. Interviews Conducted

Data Source	Fall Year 1	Winter Year 1	Spring Year 1	Total
Panel members	9		9	18
Coaches	10	3	8	21
Total	19	3	17	39

Table A3. Partial QSR NVivo Coding Schema (with relevant nodes in bold)

- (1) Making sense of new roles
 - (1 5) Distribution/sharing of role tasks and functions
 - (1 5 1) Distributed accountability
 - (1 5 1 1) Teacher union and school district
 - (1 5 1 2) **Panel and coaches**
 - (1 5 1 2 1) **panel support to coaches (reciprocity)**
 - (1 5 1 2 2) **panel holding coaches accountable**
 - (1 5 1 2 2 1) **for support of PTs**
 - (1 5 1 2 2 2) **for accountability of PTs**
 - (1 5 1 2 2 3) **ineffectively**
 - (1 5 1 3) Panel and principals
 - (1 5 1 4) Coaches and principals
 - (1 5 1 5) Principals and PTs
 - (1 7) Relationships
 - (1 7 1) Site administrator–coach
 - (1 7 2) Coach–participating teacher
 - (1 7 3) Site administrator–participating teacher
 - (1 7 4) Lead teacher–lead teacher (coach/coach or panel T/panel T)
 - (1 7 5) **Site admin–panel**
 - (1 7 6) **Coach–panel**
 - (1 7 7) Coach–lead coach
 - (1 7 8) **Panel admin–panel teacher**
- (3) Professionalizing teaching
 - (3 1) Professionalism/beliefs and values (3 1)
 - (3 1 1) **Shared knowledge base**
 - (3 1 2) Concern for client welfare
 - (3 1 3) **Collective responsibility for professional standards**
- (4) Roles (as subject)
 - (4 8) **Panel**
 - (4 8 1) **Panel hearing**
 - (4 9) **Quad**
- (9) Base codes
 - (9 1) Data source
 - (9 1 1) **Panel admin**
 - (9 1 2) **Coaches**
 - (9 1 3) Principals
 - (9 1 4) Participating teachers
 - (9 1 5) **Panel meeting**
 - (9 1 6) **Panel hearing**
 - (9 1 7) Coach meeting
 - (9 1 8) **Quad meeting**
 - (9 1 9) Archival document
 - (9 1 10) **Panel teacher**

Table A4. Research Questions Mapped Through Data Sources and Analysis

Research Questions	(1) How did the PAR panel work?	(2) How, if at all, did the presence of an oversight panel affect the teacher evaluation process?	(3) How, if at all, did the presence of an oversight panel affect personnel outcomes?
Data Sources	Observations of panel meetings and hearings, and quad meetings; interviews with panel members and coaches	Observations of panel meetings and hearings, and quad meetings; interviews with panel members and coaches	Observations of panel hearings
Data Analysis (coding node from Table A3)	(4 8) (4 8 1) (4 9) (9 1 5) (9 1 6) (9 1 8)	(1 5 1 2) (all subnodes) (1 7 5) (1 7 6) (1 7 8) (3 1 1) (3 1 3)	(1 5 1 2 2 1) (1 5 1 2 2 2) (1 5 1 2 2 3)

Notes

1. The case district presented here drew heavily from established peer assistance and review programs, in particular from Toledo, Ohio, the original model that is often used as a blueprint for other programs.

2. Although this was initially the case, Rosemont hired a new Superintendent just as PAR began, and her support for the program was fairly ambivalent. For a discussion of this and other organizational issues related to PAR implementation, see Goldstein, 2004.

3. Coach caseloads were 12–15 by contract, but they had reduced caseloads in the first year of the program to provide them with time for program development, including the creation of standards-based forms and documentation.

4. I also interviewed a sample of participating teachers (15/91) and principals (11/28) for the larger study out of which this article grew. Given this article's specific focus on the role of the PAR panel as overseer of the coaches' evaluations of classroom teachers, data sources are bounded to panel members and coaches. Readers interested in participating teachers' responses to the program should see Goldstein, 2007b, for an overview, and Goldstein, 2007a, for a discussion of the combination of formative and summative assessment, and the affect on PTs' trust in their coaches (including issues of transparency). Readers interested in the panel's relationship with principals should see Goldstein, 2007c.

5. Although such software can sometimes force researchers into analytic schema too early, I did not create my NUD*IST coding schema until all data had been collected. In addition, I created the schema from my own progressive coding patterns, unassisted by the software.

6. Different PAR locals use different terminology for the same or similar roles and concepts. "PAR coaches" are sometimes called consulting teachers, mentors, or lead teachers, for example. The PAR "panel" is sometimes a "governing board."

7. Coaches also met as a whole group all day every Friday, but coaches' professional community among themselves goes beyond the scope of this article.

8. For example, one PT in particular, who was dismissed, complained about a lack of transparency—in his words, that his coach presented herself as there to support him but

ultimately recommended his dismissal. Although interactions between this PT and his coach were not observed, reports by the coach to the panel as early as November do not support the PT's claim. More important, most of the other 14 PTs interviewed, including five who were "red flags" and one who was dismissed, reported a trusting relationship with their coach. In addition, a survey of all PTs at the end of the second year of the program (with 112 of 143 PTs responding) found that the mean for the construct "trust" in coach (3.47, $SD = 0.82$) was roughly a point and a half higher than the mean for the construct "lack of trust" in coach (1.97, $SD = 0.85$), as reported by PTs on a 4-point Likert scale (1=strongly disagree; 4=strongly agree). For a construct measuring help received, the 112 second-year PTs reported a mean of 2.97 ($SD = 0.75$) on the 4-point Likert scale. (See Goldstein, 2007a, 2007b, for reactions to PAR by PTs.)

9. For an example of a PAR program that looks much like traditional teacher evaluation, see Feiman-Nemser, 2001.

10. In Years 2–4 of the PAR program, the rate of dismissal for beginning teachers fell to 10%, still extremely high when compared with typical rates and rates, in Rosemont prior to PAR. Some believed the shift was due to fewer uncredentialed teachers being hired in the first place by the district. In addition, although the veterans placed in the program in its initial year were perceived to be notoriously below standards—the ones "we've been trying to get rid of for years"—by the third year of the program, one of the four veterans in PAR that year successfully exited the program. This still placed the district below the average of a sample of other established PAR programs, in which 30%–60% of veterans were remediated (Darling-Hammond, 1984; Kelly, 1998; Murray, 1999; Hewitt, 2000).

11. In the cases presented here, the panel challenged coaches' recommendations for renewal when the evidence did not warrant renewal. It might be expected that over a larger sample of PT cases over time, the reverse could also be true; we might expect the panel to challenge coaches who recommended dismissal when the evidence demonstrated that participating teachers were meeting standards or showing sufficient growth. Given the historic tendency in education to perpetually give satisfactory evaluations, it is not surprising that accountability for unwarranted retention (rather than unwarranted dismissal) was seen at the outset of Rosemont's PAR program.

12. It should be noted that Dal Lawrence, the former teacher union president of Toledo Public Schools and the original mastermind of peer review for K–12 teachers, argued that one person must ultimately be responsible for evaluation (D. Lawrence, personal communication, 2002). Lawrence argued that if this responsibility is spread out—and he was concerned about shared responsibility between coaches and principals—then the district opens itself to a loss in arbitration should a teacher appeal a dismissal. In other words, two or more people may disagree about performance, and any disagreement could be used to create doubt about the decision and therefore pave the way for a teacher's successful appeal. An alternate perspective, however, suggests that multiple opinions can be considered as part of a professional dialogue, and then the group has to stand behind one decision. A panel certainly needs a clear decision-making and voting process, and an agreement to stand together behind the group's decision.

References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. Chicago: University of Chicago Press.
- Baier, V. E., March, J. G., & Saetren, H. (1988). Implementation and ambiguity. In J. G. March (Ed.), *Decisions and organizations* (pp. 150–164). New York: Basil Blackwell.
- Benveniste, G. (1987). *Professionalizing the organization: Reducing bureaucracy to enhance effectiveness*. San Francisco: Jossey-Bass.
- Bridges, E. M. (1986). *The incompetent teacher: The challenge and the response*. Philadelphia: Falmer Press.
- Bridges, E. M. (1992). *The incompetent teacher: Managerial responses*. Washington, DC: Falmer Press.
- California Commission on Teacher Credentialing, State Superintendent of Instruction, and State Board of Education. (1997). *California standards for the teaching profession: A description of professional practice for California teachers*. Sacramento, CA: Author.
- Callahan, R. E. (1962). *Education and the cult of efficiency*. Chicago: University of Chicago Press.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1984). The Toledo (Ohio) public school intern and intervention programs. In A. E. Wise, L. Darling-Hammond, M. W. McLaughlin, & H. T. Bernstein (Eds.), *Case studies for teacher evaluation: A study of effective practices* (pp. 119–166). Santa Monica: RAND.
- Darling-Hammond, L. (1990). Teacher professionalism: Why and how? In A. Lieberman (Ed.), *Schools as collaborative cultures: Creating the future now* (pp. 25–50). Bristol, PA: Falmer Press.
- Darling-Hammond, L., & Sykes, G. (Eds.). (1999). *Teaching as the learning profession: Handbook of policy and practice*. San Francisco: Jossey-Bass.
- Elmore, R. (2003). *Doing the right thing, knowing the right thing to do: The problem of failing schools and performance-based accountability*. Cambridge, MA: Harvard Graduate School of Education and Consortium for Policy Research in Education.
- Elmore, R. (2006, November). *Education leadership as the practice of improvement*. Paper presented at the annual meeting of the University Council for Educational Administration, San Antonio, TX.
- Englund, T. (1996). Are professional teachers a good thing? In I. F. Goodson & A. Hargreaves (Eds.), *Teachers' professional lives* (pp. 75–87). Washington, DC: Falmer Press.
- Etzioni, A. (1969). *The semi-professions and their organization: Teachers, nurses, social workers*. New York: Free Press.
- Feiman-Nemser, S. (2001, April). *Embracing contraries: Assistance and assessment in beginning teacher induction*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Feiman-Nemser, S., & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 505–526). New York: Macmillan.
- Fideler, E. F., & Haselkorn, D. (1999). *Learning the ropes: Urban teacher induction programs and practices in the United States*. Belmont, MA: Recruiting New Teachers.
- Freidson, E. (1994). *Professionalism reborn: Theory, prophecy and policy*. Chicago: University of Chicago Press.

- Gallagher, J. J., Lanier, P., & Kerchner, C. T. (1993). Toledo and Poway: Practicing peer review. In C. T. Kerchner & J. E. Koppich (Eds.), *A union of professionals: Labor relations and educational reform* (pp. 158–176). New York: Teachers College Press.
- Goldstein, J. (2003). *Teachers at the professional threshold: Distributing leadership responsibility for teacher evaluation*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Goldstein, J. (2004). Making sense of distributed leadership: The case of peer assistance and review. *Educational Evaluation and Policy Analysis*, 26(2), 173–197. (Reprinted with apology; originally published 25(4), 397–421)
- Goldstein, J. (2007a). Debunking the fear of peer review: Combining supervision and evaluation and living to tell about it. *Journal of Personnel Evaluation in Education*, 18, 235–252.
- Goldstein, J. (2007b). Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review. *American Journal of Education*, 113, 479–508.
- Goldstein, J. (2007c). Distributed accountability: Creating district systems to ensure teaching quality. *Journal of School Leadership*, 17(4). [Special issue].
- Goldstein, J. (forthcoming). *Complexity and courage in teacher leadership: Creating professional accountability with peer assistance and review*. New York: Teachers College Press.
- Halverson, R. R. (2003). Systems of practice: How leaders use artifacts to create professional community in schools. *Education Policy Analysis Archives*, 11(37). Retrieved March 13, 2006, from <http://epaa.asu.edu/epaa/v11n37/>
- Halverson, R. R., & Zoltmers, J. (2001, April). *Distribution across artifacts: How designed artifacts illustrate school leadership practice*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Hewitt, D. (2000). The Cincinnati plan. In G. Bloom & J. Goldstein (Eds.), *The peer assistance and review reader* (pp. 115–121). Santa Cruz: New Teacher Center at the University of California, Santa Cruz.
- Hunter, M. (1988). Effecting a reconciliation between supervision and evaluation: A reply to Popham. *Journal of Personnel Evaluation in Education*, 1, 275–279.
- Kelly, P. P. (1998). *Teacher unionism and professionalism: An institutional analysis of peer review programs and the competing criteria for legitimacy*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- Kerchner, C. T., & Cauffman, K. D. (1995). Lurching toward professionalism: The saga of teacher unionism. *The Elementary School Journal*, 96, 107–122.
- Kerchner, C. T., Koppich, J. E., & Weeres, J. G. (1997). *United mind workers: Unions and teaching in the knowledge society*. San Francisco: Jossey-Bass.
- Kerchner, C. T., & Mitchell, D. E. (1988). *The changing idea of a teachers' union*. New York: Falmer Press.
- Koppich, J. E. (1998). *Spotlighting teacher quality: A review of teacher evaluation*. Paper prepared for the American Federation of Teachers, Washington, DC.
- Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal*, 19, 325–340.
- Little, J. W. (1988). Assessing the prospects for teacher leadership. In A. Lieberman (Ed.), *Building a professional culture in schools* (pp. 78–106). New York: Teachers College Press.
- Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers College Record*, 105, 913–945.
- Lortie, D. (1975). *Schoolteacher*. Chicago: University of Chicago Press.
- Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest districts. *Journal of Personnel Evaluation in Education*, 10, 203–226.
- March, J. G. (1994). *A primer on decision making: How decisions happen*. New York: Free Press.

- Miles, K. H., & Darling-Hammond, L. (1998). Rethinking the allocation of teaching resources: Some lessons from high-performing schools. *Educational Evaluation and Policy Analysis*, 20(1), 9–29.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Murphy, J., & Seashore Louis, K. (1999). *Handbook of research on educational administration* (2nd ed.). San Francisco: Jossey-Bass.
- Murray, C. E. (1999). Rochester teachers struggle to take charge of their practice. In B. Peterson & M. Charney (Eds.), *Transforming teacher unions* (pp. 46–49). Milwaukee, WI: Rethinking Schools.
- Painter, S. R. (2000). Principals' perceptions of barriers to teacher dismissal. *Journal of Personnel Evaluation in Education*, 14, 253–264.
- Rosenholtz, S. J. (1989). *Teachers' workplace*. New York: Longman.
- Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks, CA: Sage.
- Seashore Louis, K. (2006). Changing the culture of schools: Professional community, organizational learning and trust. *Journal of School Leadership*, 16, 477–489.
- Spillane, J. P. (2006). *Distributed leadership*. San Francisco: Jossey-Bass.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, 30(3), 23–28.
- Spillane, J. P., Reisen, B., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72, 387–431.
- Tucker, P. D. (1997). Lake Wobegon: Where all teachers are competent (Or, have we come to terms with the problem of incompetent teachers?). *Journal of Personnel Evaluation in Education*, 11, 103–126.
- Van Maanen, J., & Barley, S. R. (1984). Occupational communities: Culture and control in organizations. *Research in Organizational Behavior*, 6, 287–365.
- Villaraigosa, A., Strom-Martin, V., & Alquist, E. K. (1999). California Assembly Bill 1X. Introduced January 19. Amended in assembly February 12 and 18.
- Webster. (1983). *Webster's new universal unabridged dictionary*. New York: Simon and Schuster.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: RAND.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

JENNIFER GOLDSTEIN is an associate professor in the Baruch College School of Public Affairs of the City University of New York. Her areas of specialization are educational policy and leadership, and her work focuses on improving the quality of teachers and teaching in schools serving low-income children and children of color in the United States. Her recent publications include articles in the *American Educator*, the *Journal of School Leadership* and a forthcoming book on peer review from Teachers College Press.